ELSEVIER

Topology and its Applications 123 (2002) 205–234

TOPOLOGY AND ITS APPLICATIONS

www.elsevier.com/locate/topol

A combinatorial algorithm for immersed loops in surfaces ☆

J.M. Paterson

DPMMS, 16 Mill Lane, Cambridge, England, UK, CB2 1SB

Received 30 March 1999; received in revised form 31 March 2001

Abstract

In this paper, we develop a purely combinatorial algorithm which minimizes the number of double points of an immersed loop in a closed, orientable surface and converts between the ambient isotopy classes of two homotopic loops using an explicit sequence of elementary homotopies. We note that by introducing a curve-shortening flow known as the disc flow, Hass and Scott have shown that given a pair of general position, immersed loops, each with k double points, then they are homotopic through loops with at most k double points, and that this homotopy may be assumed to be regular except at finitely many points. We demonstrate this here without recourse to the geometry of the surface, by giving an explicit homotopy, which relies solely on the notion of a *spanning disc* for an immersed loop, which we define here to be an embedding of the standard 2-disc into the universal covering space of the surface, for which the boundary is mapped into the union of the lifts of the loop. The most important of these spanning discs have a 1-gon, 2-gon or 3-gon structure relative to the family of covering curves for the loop. © 2001 Published by Elsevier Science B.V.

AMS classification: 57N05; 57N10

Keywords: Surfaces; Immersions; Loops; Combinatorial

1. Introduction

In this paper, we take an arbitrary general position immersion of an oriented loop into a closed, orientable surface and present a combinatorial algorithm which homotops the loop to one with the minimal number of double points using only local homotopies of an especially simple form. We then generalize this approach by developing an algorithm which, given two homotopic immersed loops, homotops the first loop through a sequence of these elementary moves to produce a loop which is ambient isotopic to the second.

☆ Research supported by MSRI, Berkeley and Girton College, Cambridge.
E-mail address: J.Paterson@dpmms.cam.ac.uk (J.M. Paterson).

0166-8641/01/$ – see front matter © 2001 Published by Elsevier Science B.V.
PII: S0166-8641(01)00202-4

In [4], Hass and Scott gave shown, by means of techniques which relied intensively on the geometry of the surface, that it is possible to move between a pair of homotopic loops each with k double points for some $k \in \mathbb{N}$, through maps with at most k double points. In particular, they constructed a piecewise-linear analogue of curvature flow which shortened a curve by successively replacing segments with geodesics. They were not, however, able to write down an explicit algorithm which achieved this, using information purely relating to the combinatorial arrangement of self-intersections of the loop on the immersed surface.

We develop an algorithm here which does precisely this, with the added benefit of being very simple to implement. For this reason it should be more directly applicable to other combinatorial problems in two and three-dimensional manifold theory. We remark, moreover, that this procedure is intimately related to the "football-reduction" and "football-exchange" procedures, which we develop in a forthcoming paper [5].

Our approach encodes the spanning discs for a general position immersed loop into a graph which we call the *state graph* for the immersed loop. Its edges capture the relative arrangement of these spanning discs in terms of special sorts of adjacencies and inclusions. The inclusions give us a natural partial ordering, so that the "higher" up a chain a spanning sub-disc is represented, the more spanning sub-discs it has. Our techniques then hinge upon analysing the way in which performing an elementary local homotopy effects this graph.

Throughout this paper, we shall take Σ to be a closed, orientable surface other than S^2 and $c : S^1 \to \Sigma$, to be a general position immersion. We follow the convention of [3] and use c to represent both the map and its image. Writing $\widetilde{\Sigma}$ for the universal covering space of Σ, which is, in this case, a copy of $\mathbb{R}^2$, we observe that c lifts to a collection of general position immersed curves in Σ. We denote this covering family by Λ. If c is null-homotopic, then Λ consists of simple closed curves in $\widetilde{\Sigma}$. Otherwise, Λ consists of non-compact curves which we refer to as *lines*. We develop our algorithm on the basis of the following key concept, where D^2 denotes the closed 2-disk.

Definition 1.1. Suppose that Σ is a closed, orientable surface and that $c : S^1 \to \Sigma$, is a general position immersion. We shall say that an embedding,

$$e : D^2 \to \widetilde{\Sigma}, \tag{1.1}$$

is a spanning disc for c if $\partial e(D^2)$ is contained in $\bigcup \Lambda$.

In practice, we refer to the image, $\Delta = e(D^2)$, as a spanning disc for c. Note that the boundary $\partial e(D^2)$ is an embedded copy of S^1 in $\widetilde{\Sigma}$. Moreover, each such disc inherits a natural polygonal structure from $\bigcup \Lambda$. In particular, a *vertex* of Δ is a double point for Λ which is the intersection of two edges of Δ. An *edge* of Δ is a maximal (connected) segment of $\partial\Delta$, which is contained within a single curve in the covering family, Λ, and has no vertices in its interior. We shall say that a spanning disc, Δ is $\pi_1(\Sigma)$-*equivariant* if it is disjoint from $g\Delta$, for every $g \in \pi_1(\Sigma)$. A spanning disc, Δ is *innermost* if $\mathrm{Int}\Delta$ is disjoint from $\bigcup \Lambda$. We note that an innermost spanning disc need not be $\pi_1(\Sigma)$-equivariant and vice versa. In particular, a pair of innermost spanning discs, Δ and $g\Delta$, for some $g \in \pi_1(\Sigma)$, may share a common vertex. Restricting our attention to innermost, $\pi_1(\Sigma)$-

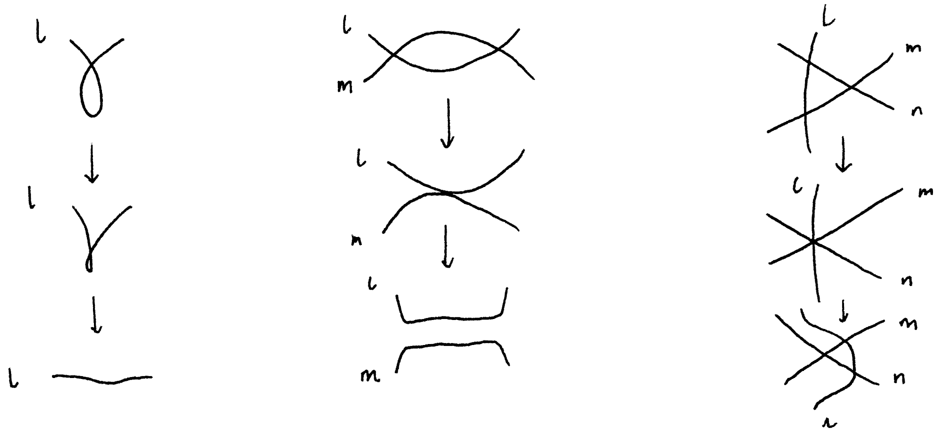

Fig. 1. Elementary homotopies.

equivariant spanning discs, we define the following three fundamental local homotopies of c.

(1) If Ξ is a $\pi_1(\Sigma)$-equivariant spanning 1-gon disc for c, then we define the 1-*gon move*, ξ, by equivariantly shrinking the disc until we eliminate it.
(2) If Γ is a $\pi_1(\Sigma)$-equivariant 2-gon disc, we define the 2-*gon move*, γ, by equivariantly deforming one edge across the other to remove the disc.
(3) If Δ is a 3-gon disc, we define the 3-*gon move*, δ, by equivariantly deforming an edge of the disc across the unique vertex outside of that edge, giving rise to a new 3-gon disc. We note that this disc may be thought of as having the opposite orientation to the original one.

We note that these moves bear some similarity to the well-known Reidemeister moves of knot theory. The key difference, however, is that these moves may always be performed across an equivariant spanning 1-gon, 2-gon or 3-gon, respectively. This is not true for Reidemeister moves which are constrained by the 3-dimensionality of the knot problem. We illustrate these in Fig. 1. Using these ideas, we prove:

Theorem 1.1. *Let $c : S^1 \to \Sigma$ be a general position immersion and Σ, a closed, orientable surface. Then the set of spanning discs for c defines a sequence of elementary moves which converts c into an immersion with the minimal number of self-intersection points in its homotopy class. Furthermore, this conversion sequence is monotonically decreasing with respect to self-intersection number.*

Theorem 1.2. *Let $c : S^1 \to \Sigma$ and $c' : S^1 \to \Sigma$ be a pair of homotopic, general position immersed loops in a closed, orientable surface, Σ. Moreover, suppose that both c and c' have precisely k self-intersection points, this being the minimal number for a general position immersion in the given homotopy class. Then the spanning discs for c and c' define a sequence of elementary moves which converts c to an immersion which is ambient* *isotopic to c'. Moreover, the curve's self-intersection number remains equal to k throughout the conversion sequence.*

2. Immersed loops in surfaces

Suppose that $c: S^1 \to \Sigma$ is a general position immersion as above with a covering family, Λ, in $\widetilde{\Sigma}$. We shall say that a pair of lines in Λ is *linked* if they meet in an odd number of points, see [1]. Suppose now that we have two homotopic, general position immersed loops, c and d. We identify the members of the two covering families of lines in $\widetilde{\Sigma}$ with a natural bijection as follows.

Definition 2.1. Suppose that $c: S^1 \to \Sigma$ and $c': S^1 \to \Sigma$ are homotopic essential immersions and h is a homotopy taking c to c' which lifts to a homotopy $\tilde{h}$, in $\widetilde{\Sigma}$ taking the line l above c to the line, l', above c'. We shall say that each line, gl, above c, where $g \in \pi_1(\Sigma)$, corresponds to gl' above c'. Using this, we define a *correspondence bijection*,

$$\Phi: \Lambda \to \Lambda' \tag{2.1}$$

$$: gl \to gl', \tag{2.2}$$

$g \in \pi_1(\Sigma)$.

We observe that a pair of linked lines in Λ corresponds under Φ to a linked pair in Λ', for any immersion, $c': S^1 \to \Sigma$ in the homotopy class of c. It is further easy to see that if a pair of lines intersect infinitely often, then they are stabilized by an infinite cyclic subgroup, $\langle g \rangle$, of $\pi_1(\Sigma)$ and the number of $\langle g \rangle$-orbits of the points of intersection is finite. We note also that the lines in Λ may have points of transverse self-intersection.

Recall from [3], that a general position immersion, c, satisfies the 1-*point intersection property* if the covering family, Λ, consists of embedded lines, any intersecting pair of which meet transversely in a single point. In [3], however, Hass and Scott constructed examples of immersed loops which cannot be homotoped to have the 1-point property. These are, in fact, curves which carry non-primitive elements of $\pi_1(\Sigma)$. A weaker property therefore has more practical value.

Definition 2.2. A general position immersion, $c: S^1 \to \Sigma$, has the minimal intersection property if it has the minimal number of self-intersection points of any general position immersed loop in its homotopy class.

The simplest example of a loop which cannot be homotoped to have the 1-point property has two lines, l and hl, in Λ, which meet transversely in an infinite number of points. These lifts cover a loop with a single point of transverse self-intersection in the cyclic covering space, Σ_h, of Σ corresponding to the sub-group, $\langle h \rangle$, of $\pi_1(\Sigma)$. We may generalize this to a situation involving an n-tuple of distinct lines, $(l, hl, h^2 l, \dots, h^{n-1} l)$, for which $|l \cap h^k l| = \infty$, $k = 1, \dots, n-1$, and $h^n l = l$, $n \geqslant 2$. If, in addition, there are two $\langle h \rangle$-orbits of these intersection points for each k where $n \geqslant 3$ and one if $n = 2$, then standard cut-and-paste arguments tell us that this is the minimal number of transverse intersections. We note that a feature such as this in Λ occurs precisely when c winds n times around a primitive loop in Σ. We refer to a feature of this form as an *n-strand* in Λ. If an n-strand is disjoint from its translates, then the primitive loop around which c winds may be embedded in the surface Σ. We review next some important definitions from [3].

Definition 2.3. If $c : S^1 \to \Sigma$ is a general position immersion, then c has a *singular* 1-gon if there is a sub-arc, α, of S^1 such that the map c identifies the endpoints of α and $c|_\alpha$ is a null-homotopic loop in Σ. Moreover, a disc D is an embedded 1-gon for c if there exists a sub-arc, α, of S^1 with $c(\alpha) = \partial D$ and the map $c|_\alpha$ injects. Similarly, c has a *singular* 2-gon if there exists disjoint sub-arcs, α and β, of S^1 such that c identifies each endpoint of α to a distinct endpoint of β and $c|_{\alpha\cup\beta}$ is a null-homotopic loop in Σ. Moreover, a disc, D, is an *embedded* 2-*gon* for c if there exist disjoint sub-arcs, α and of β, S^1 which embed under c, where $c(\alpha) \cup c(\beta) = \partial D$ and $c(\alpha) \cap c(\beta) = c(\partial\alpha) = c(\partial\beta)$. Finally, c has a *singular* 3-*gon* if there exist disjoint sub-arcs, α, β and γ, of S^1 such that c cyclically identifies their endpoints and $c|_{\alpha\cup\beta\cup\gamma}$ is a null-homotopic loop in Σ. Moreover, a disc D is an *embedded* 3-*gon* for c if there exist disjoint sub-arcs, α, β and γ of S^1 which embed under c for which $c(\alpha) \cup c(\beta) \cup c(\gamma) = \partial D$ and each of the sets, $c(\alpha) \cap c(\beta)$, $c(\beta) \cap c(\gamma)$ and $c(\alpha) \cap c(\gamma)$ consists of a single point.

Remark 1. If c has an embedded k-gon, D, $k = 1, 2, 3$, then D lifts to a $\pi_1(\Sigma)$-equivariant spanning k-gon for c in $\widetilde{\Sigma}$.

We next paraphrase Theorems 4.2 and 2.7 of [3] as follows:

Theorem 2.1 (Hass and Scott). *Suppose that c is a general position immersion of S^1 into the closed orientable surface Σ. If c does not have the minimal intersection property, then it has a singular* 1*-gon or* 2*-gon. Furthermore, if c is homotopic to an embedded loop, then c has an embedded* 1*-gon or* 2*-gon.*

It follows from this that any null-homotopic loop in Σ has a $\pi_1(\Sigma)$-equivariant spanning 1-gon or 2-gon across which an elementary move may be performed, thereby reducing the number of self-intersection points. Given a singular loop which is homotopic to an embedding, this gives us a natural procedure for achieving an embedded loop via a sequence of 1-gon and 2-gon moves. Our chief interest, therefore, lies with essential loops which are not homotopic to embeddings. We note that in [3], Hass and Scott give some explicit constructions of immersed loops without the minimal intersection property which have no embedded 1-gon or 2-gon discs.

Lemma 2.1. *Suppose that $c : S^1 \to \Sigma$ is a general position, essential immersion and Σ is a closed, orientable surface. If c has a spanning* 1*-gon,* 2*-gon or* 3*-gon, then there exists an innermost spanning n-gon for some $n \leqslant 3$.*

Proof. By a traversing segment of an n-gon disc, we mean an immersion of the closed unit interval into the n-gon where the endpoints are mapped into the boundary. Using this, we proceed case-wise.

(1) *The loop c has a spanning* 1*-gon*: We choose a sub-1-gon which is small. It follows that any traversing segment must be embedded. Then either the 1-gon is innermost and the conclusion is immediate or it has a sub-2-gon. In the latter case we have the situation of 2 below.

(2) *The loop c has a spanning* 2*-gon*: We choose a sub-2-gon which is small. If this disc is non-innermost, then it has either a singular or an embedded traversing segment. In the former case, it has a sub-1-gon and we argue as in 1 above. In the latter case, we have a sub-3-gon and we deal with this in 3 below.

(3) *The loop c has a spanning* 3*-gon*: We choose a sub-3-gon which is small. If this disc is non-innermost, we consider, as in 2, the possibilities of singular or embedded traversing segments. In the event of the former, we have a sub-1-gon and argue as in case 1. In the event of the latter there must be a sub-2-gon and we argue as in 2.

By compactness of the original n-gon, this process must terminate with a spanning disc which is a 1-gon, 2-gon or 3-gon, as claimed. □

Suppose, in the sequel, that c is an oriented loop. We define the *positive* side of c to be the side for which the normal to the curve is inward-pointing. We label the closed half-planes defined by a line, l_s, in Λ by l_s^{+1} and l_s^{-1}, where the superscripts are consistent with our orientation on c. Using this, the spanning discs for c may be expressed naturally as components of the intersection of closed half-planes in $\widetilde{\Sigma}$.

Definition 2.4. Suppose that Δ is a spanning n-gon disc for c with boundary segments in the n-tuple of lines, $(l_1, l_2, \ldots, l_n)$ in Λ, where two or more of these components may coincide. We shall say that the ordered n-tuple, $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ in $\{-1, +1\}^n$ is the $(l_1, l_2, \ldots, l_n)$-*intersection index* of Δ if Δ is a component of

$$\bigcap_{i=1}^{n} l_i^{\varepsilon_i}. \tag{2.3}$$

If c has the 1-point property, then we shall say that $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is the *intersection index* for the n-tuple $(l_1, l_2, \ldots, l_n)$.

We note that it is often convenient, when the n-tuple of lines is clear from the context, to refer to the *intersection index* of a spanning n-gon disc. In particular, we fix an ordering of these n lines for any given disc to ensure uniqueness. We note that intersection index is a $\pi_1(\Sigma)$-equivariant quantity, that is it is invariant under the action of $\pi_1(\Sigma)$.

Remark 2. If c has the 1-point property, then the intersection index of a triple of mutually intersecting lines is invariant under transverse homotopy up to negation in each component. This corresponds to the two possible configurations for a spanning 3-gon with boundary segments in these lines.

We note that a 3-gon move produces a 3-gon disc with intersection index the negative of that of the original disc.

Suppose next that c and c' are homotopic essential immersions. We would like to compare the spanning discs for c with those of c'. In order to do this, we attempt to pair spanning discs for c and c' in a natural way.

Definition 2.5. Suppose that Δ is a spanning n-gon for c which is a component of $\bigcap_{i=1}^{n} l_i^{\varepsilon_i}$. We shall say that Δ properly corresponds to a spanning n-gon, Δ', for c' if Δ' is a component of $\bigcap_{i=1}^{n} l_i'^{\varepsilon_i}$, where l_i' corresponds to l_i. If Δ is a 3-gon in $\bigcap_{i=1}^{3} l_i^{\varepsilon_i}$, then we shall say that Δ reverse corresponds to Δ' if Δ' is a component of $\bigcap_{i=1}^{3} l_i'^{-\varepsilon_i}$.

We note that if c has the 1-point property, then this relation of correspondence defines a bijection between the spanning discs for c and those for c'. In general, however, this is not the case and attempts to define an injective correspondence leave us with 1-gon, 2-gon or 3-gon discs in $\widetilde{\Sigma}$ with no counterparts in $\widetilde{\Sigma}'$ and vice versa. Likewise, a 1-gon, 2-gon or 3-gon disc for c may have a number of corresponding discs in $\widetilde{\Sigma}$. We note, moreover, that if this relation is bijective on the 1-gons, 2-gons and 3-gons and all the correspondences of 3-gons are proper, then c and c' are ambient isotopic. Attempting to construct a bijection in this way to decide whether two immersions are ambient isotopic appears at first to present a rather daunting task. By encoding the information of inclusions and adjacencies of spanning discs into a graph, the problem reduces to recognizing isomorphisms of graphs. This graph has the practical advantage of determining all the explicit sequences of elementary moves which change the ambient isotopy class of an immersed loop. In order to build this graph, we start with the following definitions.

Definition 2.6. We shall say that a pair of spanning discs, Δ and Γ, are *vertex adjacent* if they have opposite angles at some common vertex. Likewise, we shall say that a pair of spanning discs are *edge adjacent* if they locally oppose across some common boundary edge sub-segment. A spanning disc, Δ, *vertex includes* a spanning disc, Γ, if Δ contains Γ and the two discs share a common vertex. A disc, Δ, *edge includes* Γ if the Δ contains Γ and they share a common edge sub-segment of their boundaries but Δ does not vertex include Γ. We shall refer to inclusion which is neither of the vertex or edge type as *interior inclusion*.

We note that where a disk includes another disk it cannot be adjacent to that disk and vice versa. We further note that a pair of 1-gons cannot be vertex adjacent if the loop is essential but that they may be edge adjacent. We illustrate these inclusions and adjacencies with examples in Fig. 2 below.

Definition 2.7. Let $c : S^1 \to \Sigma$ be an oriented immersion. We define a graph, $\widetilde{S}(c)$, of c the vertices of which are the 1-gon, 2-gon and 3-gon spanning discs for c in $\widetilde{\Sigma}$ and the

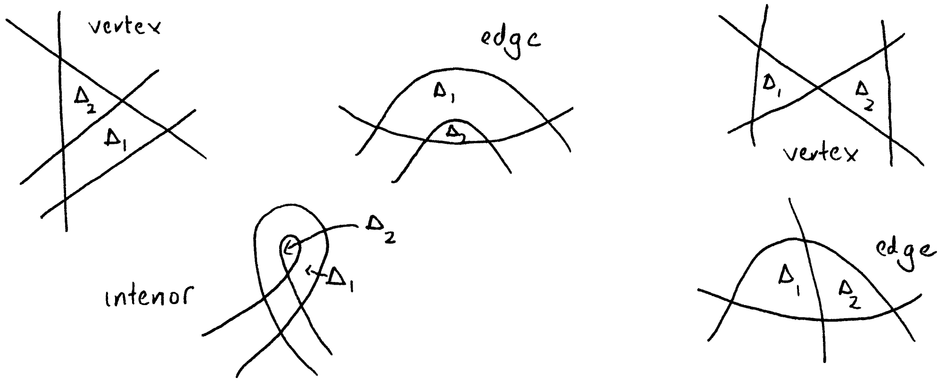

Fig. 2. Inclusions and adjacencies.

edges of which are defined as follows. If Δ_1 and Δ_2 are vertex adjacent spanning discs for c, then we write

$$\Delta_1 \quad \longleftrightarrow \quad \Delta_2. \tag{H1}$$

If Δ_1 and Δ_2 are edge adjacent and $\Delta = \Delta_1 \cup \Delta_2$, then we write

$$\Delta_1 \quad \Longleftrightarrow \quad \Delta_2. \tag{H2}$$

We say, moreover, that Δ_1 and Δ_2 are *paired* in Δ, and that Δ_1 is the *partner* of Δ_2 in Δ. If Δ_1 vertex includes Δ_2, then we write

$$\begin{array}{c}\Delta_1\\ \downarrow\\ \Delta_2.\end{array} \tag{V1}$$

If Δ_2 is a 2-gon which is edge included by Δ_1, then we also define an edge of the form (V1) above. Similarly, if Δ_2 is a 1-gon and the inclusion is interior inclusion. Finally, if Δ_1 is a 2-gon, Δ_2 is a 1-gon and the pair show the special case of vertex inclusion shown in Fig. 3, then we define an edge

$$\begin{array}{c}\Delta_1\\ \Downarrow\\ \Delta_2.\end{array} \tag{V2}$$

We then define the *state graph*, $\boldsymbol{S}(c)$, to be the quotient graph of $\widetilde{\boldsymbol{S}}(c)$ under the action of $\pi_1(\Sigma)$. Hence the vertices of $\boldsymbol{S}(c)$ are the $\pi_1(\Sigma)$-orbits of the spanning 1-gon. 2-gon and 3-gon discs for c. Representing these vertices by a collection, $\boldsymbol{D}(c)$, of spanning discs, we may assign the superscript (g), $g \in \pi_1(\Sigma)$, to a vertical edge of $\boldsymbol{S}(c)$ if the edge results from an inclusion of $g.\ \Delta_2$ by Δ_1 in $\widetilde{\boldsymbol{S}}(c)$, where Δ_1 and Δ_2 are discs in $\boldsymbol{D}(c)$. Similarly, we may add a superscript $(g) \rightarrow$ to a horizontal edge resulting from the adjacency of Δ_1 and $g\Delta_2$ in $\widetilde{\boldsymbol{S}}(c)$, $\Delta_1, \Delta_2 \in \boldsymbol{D}(c)$. We refer to the edges of resulting from type (H1) and (H2) edges in $\widetilde{\boldsymbol{S}}(c)$ as *horizontal* edges in $\boldsymbol{S}(c)$ and those of type (V1) and (V2) as *vertical* edges in $\boldsymbol{S}(c)$. If the group element on an edge is the identity for our choice of $\boldsymbol{D}(c)$, we omit it.

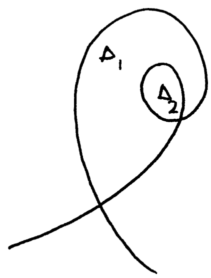

Fig. 3. A (V2) vertex inclusion.

We note that inclusions other than those of type (V1) and (V2) above are omitted from $\boldsymbol{S}(c)$. In general, when representing sub-graphs of $\boldsymbol{S}(c)$, we choose $\boldsymbol{D}(c)$ so that some vertical or horizontal edge carries the identity element in $\pi_1(\Sigma)$. Suppose that we have constructed $\boldsymbol{S}(c)$ in the above way. We observe that a type (H1) edge may connect a vertex, v_i, to itself, although orientation arguments prevent similar behaviour with a type (H2) edge. We denote this feature by $\boldsymbol{v}_i^{(g)}$, where the superscript (g) refers to an edge of the form $\overset{(g)\rightarrow}{\longleftrightarrow}$. We refer to $\boldsymbol{v}_i^{(g)}$ as a *unit circuit* on v_i. We note that a unit circuit on a 2-gon type vertex occurs precisely when we have a pair of lifts, l and gl, of c which meet transversely in an infinite number of points, all of which project to a single point in the covering space corresponding to the infinite cyclic group, $\langle g\rangle$. We further note that a unit circuit on a 3-gon type vertex has a representative disc, Δ, with boundary segments in lines of the form l, gl and g^2l, where $g \in \pi_1(\Sigma)$ and the (l, gl, g^2l)-intersection index of Δ is either $(1, -1, 1)$ or $(-1, 1, -1)$. It is easy to see that a 1-gon vertex cannot support a unit circuit. We note, moreover, that a 3-gon unit circuit gives us the sole example of a non-equivariant, yet innermost, spanning 3-gon.

Remark 3. The graph $\boldsymbol{S}(c)$ is uniquely determined by the ambient isotopy class of c up to the $\pi_1(\Sigma)$ superscripts which vary according to the choice of $\boldsymbol{D}(c)$. An alternative viewpoint considers $\boldsymbol{S}(c)$ as the quotient of a graph whose vertices are all the spanning discs for c by the action of $\pi_1(\Sigma)$.

Lemma 2.2. *The state graph, $\boldsymbol{S}(c)$, of a general position, essential immersion, $c : S^1 \to \Sigma$ is a finite graph.*

Proof. Since c has only finitely many double points, there must exist maximal spanning 1-gon and 2-gon discs with respect to inclusion. Furthermore, there must exist innermost 1-gon, 2-gon and 3-gon discs as Σ is compact, see [3]. It remains, however, to exclude the possibility of an infinite sequence of nested 3-gons. We start by considering the case where the genus of Σ is at least two. Then we may equip $\widetilde{\Sigma}$ with a hyperbolic metric. Given the finite number of double points of c, the infinite collection of lines must partition into three finite collections, together with their translates by three infinite cyclic groups. These cyclic groups are the stabilizer sub-groups of the three lines which bound the smallest 3-gon in the chain of inclusions. These groups, however, must be one and the same in order to allow this infinite nesting. To see this, we observe that the members $\pi_1(\Sigma)$ act as loxodromic actions in the hyperbolic plane from which it follows easily that the configuration is impossible. It therefore remains to consider the case where Σ is a torus. Even then, however, the configuration in question cannot result from lifts of single curve but requires instead three curves where each has lifts in precisely one of the three families defined above. $\square$

We next describe some of the important features of $\boldsymbol{S}(c)$.

Definition 2.8. A vertical chain in $\boldsymbol{S}(c)$ headed by a vertex, v_i, in $\boldsymbol{S}(c)$, is a sub-graph consisting of vertices, $v_{i_0} = v_i,\ v_{i_1}, \ldots, v_{i_n}$ and vertical edges, such that each v_{i_k} lies beneath $v_{i_{k-1}}$, $k \in \mathbb{N}$. The shadow, s_i, of a vertex, v_i, in $\boldsymbol{S}(c)$ is the full sub-graph of $\boldsymbol{S}(c)$ on the union of all vertical chains headed by v_i.

A vertex v_i in $\boldsymbol{S}(c)$ is *minimal* if there is no vertex v_j such that a vertical edge runs from v_i to v_j. For example, given a 3-gon Δ, any segment which crosses Δ parallel to an edge of Δ produces a new 3-gon, Δ_{i_1}, which is (V1)-included by Δ. Hence the vertex v corresponding to Δ cannot be minimal in $\boldsymbol{S}(c)$. The vertical edges give us a natural partial ordering on the vertices of $\boldsymbol{S}(c)$. In particular, a minimal vertex in $\boldsymbol{S}(c)$ with respect to this ordering is the $\pi_1(\Sigma)$-orbit of an innermost spanning disc for c. We note, moreover, that the vertices in a vertical chain in $\boldsymbol{S}(c)$ may be represented by nested spanning discs for c.

Example 2.1. We illustrate a possible local arrangement of spanning discs in $\widetilde{\Sigma}$, in Fig. 4, together with the associated state sub-graph. The spanning discs may be described by the following counter-clockwise vertex triples: Δ_{i_1}: (P_1, P_2, P_3), Δ_{i_2}: (P_1, P_5, P_6), Δ_{i_3}: (P_1, P_4, P_7), Δ_{i_4}: $(g.P_1, P_4, P_4)$, Δ_{i_5}: $(g.P_1, P_7, P_6)$.

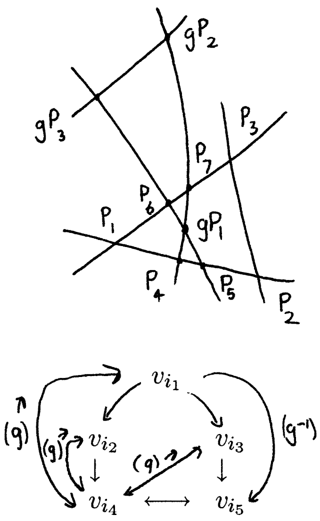

Fig. 4.

Definition 2.9. We shall say that a sub-graph of $\boldsymbol{S}(c)$ with vertex set $\{v_{i_1}, \ldots, v_{i_n}\}$ is a *length n vertical circuit* if there exists a choice for $\boldsymbol{D}(c)$ which produces a sub-graph with one of the following n configurations:

$$\begin{tikzcd} v_{i_1} \arrow[d] \\ \vdots \arrow[d] \\ v_{i_n}, \arrow[uu, bend left=60, "(g)"] \end{tikzcd} \tag{2.4}$$

which we call a *standard form vertical circuit*, or one of the form

$$\begin{tikzcd} v_{i_1} \arrow[dd] \arrow[dr, bend left, "(g)"] & \\ & v_{i_n} \\ \vdots \arrow[d] & \vdots \arrow[d] \\ v_{i_{n-k}} \arrow[r, leftrightarrow] & v_{i_{n-k+1}}, \end{tikzcd}$$

where $k \in \{1, \ldots, n-1\}$. We denote the standard vertical circuit, 2.4, by $\boldsymbol{v}^{g}_{i_1, i_n}$. In each of these situations, we shall say that the vertex v_{i_1} *heads* the vertical circuit.

An example of a vertical circuit in which the turning vertex is a 2-gon is shown in Fig. 5, where the 2-gon Γ has vertices gP_2 and Q, the 3-gon Δ has vertices (P_1, P_2, P_3) and the 3-gon Δ_1 has vertices (P_4, P_2, P_5).

We leave it as an easy exercise for the reader to check that if a spanning disc, Δ, meets a non-trivial translate, $g\Delta$, then the corresponding vertex for Δ in $\boldsymbol{S}(c)$ heads a vertical circuit.

Another important class of sub-graphs of $\boldsymbol{S}(c)$ is described below.

Definition 2.10. We shall say that a sub-graph of $\boldsymbol{S}(c)$ is a lateral escape sub-graph with *turning vertex v_i* if it has one of the following forms:

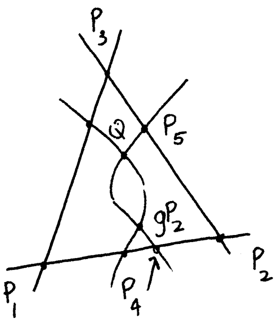

Fig. 5.

LE(A):
$$\begin{tikzcd} v_u \arrow[d] & & v_r \arrow[d, "(g)"] \\ v_i \arrow[rr, leftrightarrow, "(h)\rightarrow"] & & v_s, \end{tikzcd} \tag{2.5}$$
where each of the vertices in this sub-graph are 3-gon type vertices and the following conditions are satisfied:
(1) v_s is (H1)-related to the partner of v_i with respect to v_u,
(2) v_i is (H1)-related to the partner of v_s with respect to v_r.

LE(B): a sub-graph of the form (2.5), where v_u and v_i are 2-gon type vertices and v_s and v_r are 3-gon type vertices and we have conditions 1 and 2 above, or a sub-graph of the form
$$\begin{tikzcd} v_u \arrow[d] & & \\ v_i \arrow[rr, leftrightarrow, "(g)\rightarrow"] & & v_s, \end{tikzcd} \tag{2.6}$$
where v_u and v_s are 3-gons and v_u is a 2-gon vertex. Moreover v_s is edge adjacent to v_u. A special sort of LE(B) lateral escape sub-graph is known as an LE(B)* lateral escape. In this case each of the vertices of the 2-gon Γ_i supports a vertex adjacency with a type LE(B) lateral escape 3-gon.

LE(C): a sub-graph of the form (2.5), where v_u and v_r are 2-gon type vertices and v_i and v_s are 3-gon type vertices and we have conditions 1 and 2 above.

LE(D):
$$v_{s_1} \longleftrightarrow v_{s_2} \longleftrightarrow \cdots \longleftrightarrow v_i \longleftrightarrow \cdots \longleftrightarrow v_{s_{2k+1}}, \tag{2.7}$$
where $k \in \mathbb{N}$ and each of the vertices $v_{s_1}, \ldots, v_{s_{2k+1}}$, is a 2-gon vertex and neither v_{s_1} nor v_{s_2} is (H1)-related to another 2-gon outside this sub-graph.

LE(E):
$$\begin{tikzcd} & v_u \arrow[dl, bend right] \arrow[dr, "(g)"] & \\ v_i \arrow[rr, leftrightarrow, "(g)\rightarrow"] & & v_s, \end{tikzcd} \tag{2.8}$$
where either v_u is a 2-gon type vertex and v_i and v_u are 3-gon type vertices or v_u is a 1-gon vertex split into a 2-gon and a 3-gon vertex.

We refer to each of the vertices, v_s, and each of the 2-gon vertices, $v_{s_1}, \ldots, v_{s_{2k+1}}$, in LE(D) above as lateral escape vertices.

In the argument that follows, these lateral escape sub-graphs play the significant role of allowing us to perform a local homotopy for some 1-gon, 2-gon or 3-gon spanning disc in the initial presence of a non-equivariant sub-2-gon or sub-3-gon. In order to explain better what is meant by this, we first discuss the geometric pictures described by the lateral escape sub-graphs, LE(A)–LE(E), above. In particular, we claim that these configurations are precisely those shown in Figs. 6, 7 and 8. That these are the configurations determined by the sub-graphs follows immediately in cases LE(D) and LE(E). In cases LE(A), LE(B) and LE(C), it is easy to see that the conditions 1 and 2 exclude all other configurations. Inspecting the LE(A) picture it is clear that we may reverse the 3-gon Δ_u without reversing Δ_i, by first reversing the 3-gon Δ_s followed by Δ_r and finally Δ_u. In the LE(B) case,

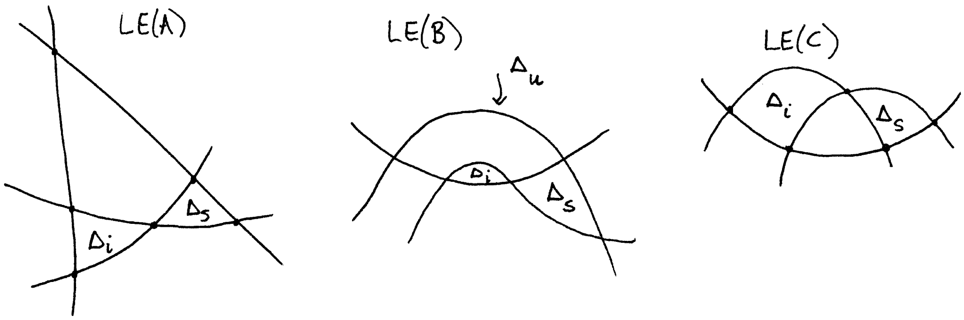

Fig. 6. Lateral escapes of types LE(A), LE(B) and LE(C).

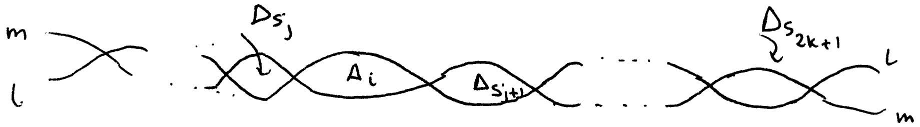

Fig. 7. A lateral escape of type LE(D).

Fig. 8. Lateral escapes of type LE(E).

we can similarly remove the large 2-gon, Δ_u, without first removing Δ_i. In the LE(C) situation, we can remove the 2-gon Δ_u, without first reversing the 3-gon Δ_i, and in the LE(D) case, performing $k+1$ 2-gon moves removes the 2-gon disc, Δ_i. In the LE(E) case, we are able to remove Δ_u by acting first on Δ_s instead of Δ_i, should Δ_i prove intractable.

One instance when we might look for lateral escape sub-graphs is when the vertex v_i either supports a unit circuit or heads a vertical circuit in $\boldsymbol{S}(c)$. We also use them to determine how great an obstacle a vertical circuit poses.

Definition 2.11. We shall say that a vertical circuit, $\boldsymbol{v}^{(g)}_{i_1,i_2}$, is *genuine* if none of the vertices $\mathrm{inv}\{v_{i_2},\ldots,v_{i_n}\}$ is the turning vertex for either a type LE(D) sub-graph or a lateral escape sub-graph of type LE(A), LE(B), LE(C) or LE(E) in $\boldsymbol{S}(c)$, itself containing a sub-graph of the form

$$\begin{array}{c} v_{i_{k-1}} \\ \downarrow \\ v_{i_k} \end{array} \tag{2.9}$$

or

$$\begin{array}{c} v_{i_1} \\ \downarrow \\ v_{i_n}. \end{array} \tag{2.10}$$

We note that if a vertex, v_i, either supports a unit circuit or heads a genuine vertical circuit, then it is impossible to act on this vertex by the obvious transverse local move. In particular, if c has the 1-point property and Γ_i is a 3-gon, then any general position immersion, $c' : S^1 \to \Sigma$ in the same homotopy class as c must have a corresponding unit or vertical circuit. A non-genuine vertical circuit, however, may be "broken" by reversing the lateral escape vertex first. We illustrate an example of this in Fig. 9. Bearing all these observations in mind, we attempt to define, given a vertex of $\boldsymbol{S}(c)$, a sub-graph which incorporates all these features and defines a natural sequence of elementary moves which allow subsequent action on v_i. We draw the readers attention to the fact that this construction is the natural ancestor of the 3-dimensional namesake, introduced in [5] for immersed π_1-injective surfaces in 3-manifolds.

Definition 2.12. Let v_i be a vertex in $\boldsymbol{S}(c)$. *A lateral escape structure*, $\boldsymbol{S}(v_i)$, for v_i is a finite a sub-graph of $\boldsymbol{S}(c)$, which we define by induction on a quantity which we call the *order* of $\boldsymbol{S}(v_i)$.

We start by defining the structure sub-graphs of order one to be the minimal vertices in $\boldsymbol{S}(c)$ which do not support unit circuits.

Suppose that we have defined structure sub-graphs of all orders less than k. Then $\boldsymbol{T}(v_i)$ is a structure sub-graph of order k if it has the following properties:

If v_i is the turning vertex for a type LE(D) lateral escape sub-graph in $\boldsymbol{S}(c)$ with $2s+1$ lateral escape 2-gons, $s \geqslant 0$, then $\boldsymbol{T}(v_i)$ contains this lateral escape sub-graph, together with structure sub-graphs of order less than k for precisely $s+1$ of the lateral escape

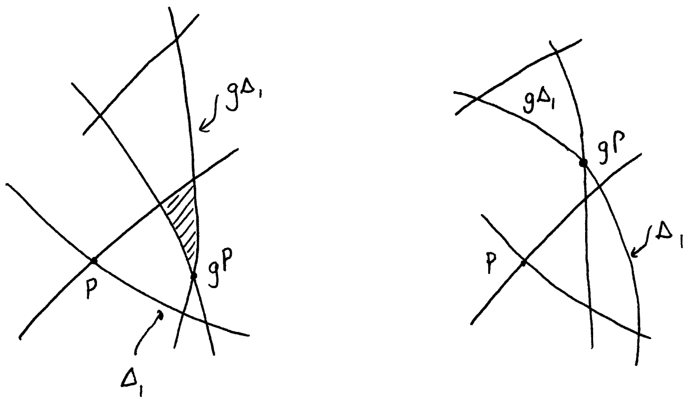

Fig. 9. Breaking a vertical circuit.

2-gons, no two of these being (H1) related. If, on the other hand, v_i is not a turning vertex of this sort, we consider the set of all sub-graphs of the form:

$$\begin{tikzcd} & v_i \arrow[dl] \arrow[dr, "(g)"] & \\ v_{i_1} \arrow[rr, leftrightarrow, "(g)\to"] & & v_{i_2}. \end{tikzcd} \tag{2.11}$$

We deal with the possibilities as follows:

(1) If v_i is a 2-gon we have one of the following:

(a) If v_{i_1} is a 2-gon, then $\boldsymbol{T}(v_i)$ contains one of

(i)
$$\begin{array}{c} v_i \\ \downarrow \\ v_{i_1}, \end{array} \tag{2.12}$$

joined with a structure sub-graph, $\boldsymbol{T}(v_{i_1})$, for v_{i_1} of order less than k or

(ii) there exists a type LE(B) lateral escape sub-graph

$$\begin{tikzcd} v_{i_0} \arrow[d] & v_{s_0} \arrow[d, "(g)"] \\ v_{i_1} \arrow[r, "(h)"] & v_{s_1}, \end{tikzcd} \tag{2.13}$$

joined with a structure sub-graph, $\boldsymbol{T}(v_0)$, of order less than k for v_{s_0} containing the vertex v_{s_1}.

(b) If v_{i_1} is a 3-gon, then $\boldsymbol{T}(v_i)$ contains one of

(i) the sub-graph (2.12), together with a structure sub-graph for v_{i_1} of order less than k,

(ii) the sub-graph

$$\begin{array}{c} v_i \\ \downarrow^{(g)} \\ v_{i_2}, \end{array} \tag{2.14}$$

together with a structure sub-graph for v_{i_2} of order less than k,

(iii) a type LE(C) lateral escape sub-graph, (2.13), joined with a structure sub-graph, $\boldsymbol{T}(v_0)$, for v_{s_0} containing the vertex v_{s_1} of order less than k,

(iv) a type LE(A) lateral escape sub-graph,

$$\begin{tikzcd} v_i \arrow[d] & \\ v_{i_0} \arrow[d] & v_{s_0} \arrow[d, "(g)"] \\ v_{i_1} \arrow[r, "(h)"] & v_{s_1}, \end{tikzcd} \tag{2.15}$$

joined with structure sub-graphs, $\boldsymbol{T}(v_{i_0})$ and $\boldsymbol{T}(v_{s_0})$, of orders less than k.

(v) If v_{i_1} is a 1-gon, then $\boldsymbol{T}(v_i)$ contains (2.12), together with a structure sub-graph for v_{i_1} of order less than k.

(c) If v_i is a 1-gon, we have the following possibilities:
   (i) v_{i_1} is a 1-gon and $\boldsymbol{T}(v_i)$ contains the sub-graph (2.12), together with a structure sub-graph for v_{i_1} of order less than k,
   (ii) v_{i_1} is a 2-gon and $\boldsymbol{T}(v_i)$ contains the sub-graph (2.12), together with a structure sub-graph for v_{i_1} of order less than k,
   (iii) v_{i_2} is a 3-gon and $\boldsymbol{T}(v_i)$ contains the sub-graph (2.12), together with a structure sub-graph for v_{i_1} of order less than k,
   (iv) any of (i), (ii) or (iii) above, replacing v_{i_1} by v_{i_2}.

(d) If v_i is a 3-gon, we have the following possibilities:
   (i) If v_{i_1} is a 2-gon, then either $\boldsymbol{T}(v_i)$ contains (2.12), together with a structure sub-graph for v_{i_1} of order less than k or $\boldsymbol{T}(v_i)$ contains a type LE(B) lateral escape sub-graph and lateral escape structures as in 1(a)(ii) above.
   (ii) If v_{i_1} is a 3-gon, then either $\boldsymbol{T}(v_i)$ contains (2.12), together with a structure sub-graph for v_{i_1} of order less than k or $\boldsymbol{T}(v_i)$ contains a type LE(A) lateral escape sub-graph, (2.13) and lateral escape structures as in 1(b)(iii).

(2) If v_i is not a minimal vertex in $\boldsymbol{T}(v_i)$ and either heads a vertical circuit or supports a unit circuit, then we have one of the following:
   (a) The vertex v_i heads a non-genuine vertical circuit, $\boldsymbol{v}_{i,i_n}^{(g)}$ in which case the vertices in $\boldsymbol{v}_{i,i_n}^{(g)}$, and the lateral escape sub-graph are contained in $\boldsymbol{T}(v_i)$, together with structure sub-graphs of order less than k as in 1 above,
   (b) v_i heads a genuine vertical circuit, $\boldsymbol{v}_{i,i_n}^{(g)}$ and v_i is the turning vertex in a type LE(D) lateral escape sub-graph which is contained in $\boldsymbol{T}(v_i)$ together with structure sub-graphs for $2s-1$ of the $2s+1$ lateral escape 2-gon vertices of order less than k.

(3) The vertices of $\boldsymbol{T}(v_i)$ are *compatible*, that is, no pair of non-turning 3-gon vertices in the above construction are horizontally related in $\boldsymbol{S}(c)$.

We refer to the set of all turning vertices in the above constructions as the *turning vertex set* for $\boldsymbol{T}(v_i)$ and the set of all lateral escape vertices as the *lateral escape vertex set* for $\boldsymbol{T}(v_i)$.

If v_i has a lateral escape structure of order k for some $k \in \mathbb{N}$, then we shall say that v_i is *weak*. If this is not the case, then v_i is *strong* in $\boldsymbol{S}(c)$.

Remark 4. We note that the 2-gons which result when a curve winds several times around a primitive loop with the minimal number of transverse intersections are all strong.

Occasionally, we need a notion of compatibility for a pair of lateral escape structures for different vertices in $\boldsymbol{S}(c)$. This situation arises, for instance, when we attempt to eliminate two different 2-gons using sequences of moves based on lateral escape structures as in Section 3 below. To this end, we say that a lateral escape structure for v_i in $\boldsymbol{S}(c)$ is *compatible* with a lateral escape structure for a different vertex, v_j, in $\boldsymbol{S}(c)$ if there exists no lateral escape, non-turning 3-gon vertex in the former which is horizontally related to a lateral escape, non-turning vertex in the latter.

3. Action on lateral escape structures

We next use lateral escape structures to define sequences of 1-gon, 2-gon and 3-gon moves which either remove a 1-gon or weak 2-gon or equivariantly reverse a weak 3-gon. We start by defining a complexity, $C(v_i)$, for the vertex v_i to be the number of distinct vertices in the lateral escape structure $S(v_i)$. Our sequence is constructed so as to decrease $C(v_i)$, noting that $C(v_i) = 0$ occurs only if v_i is minimal in $\boldsymbol{S}(c)$. In order to do this, we first examine the local effects of these generic moves on $\boldsymbol{S}(c)$. To start with, suppose that v_j is a 3-gon vertex which is minimal and does not support a unit circuit in $\boldsymbol{S}(c)$. Writing v_j^{-1} for the 3-gon vertex replacing v_j after performing the 3-gon move, δ_j, we have

$$\begin{array}{c} v_i \\ \downarrow{\scriptstyle (g)} \\ v_j \end{array} \xrightarrow{\delta_j} v_i \xleftrightarrow{(g)\rightarrow} v_j^{-1} \tag{3.1}$$

and conversely,

$$v_i \xleftrightarrow{(g)\rightarrow} v_j \xrightarrow{\delta_j} \begin{array}{c} v_i \\ \downarrow{\scriptstyle (g)} \\ v_j^{-1}. \end{array} \tag{3.2}$$

In one very special instance, namely that of a (V2)-inclusion, a 2-gon is converted to a 1-gon by performing the 1-gon move for the smaller vertex, see Fig. 10.

We next examine the effect of these moves on lateral escape sub-graphs. In particular, if v_j is the lateral escape vertex in a type LE(A) sub-graph,

$$\begin{tikzcd} v_u \arrow[d] & v_r \arrow[d, "(g)"] \\ v_i \arrow[r, leftrightarrow, "(h)\rightarrow"] & v_j \end{tikzcd} \tag{3.3}$$

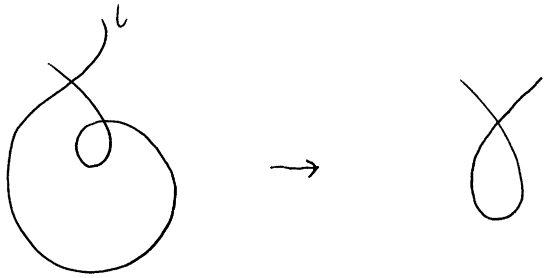

Fig. 10.

and v_j is minimal in $\boldsymbol{S}(c)$, and does not support a unit circuit, then we have

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \vdots \\
v_u \arrow[d] & v_r \arrow[d, "(g)"] & v_i \arrow[d, "(g)"] & v_u \arrow[d, "(g^{-1}h)"] \\
v_i \arrow[r, leftrightarrow, "(h)\rightarrow"] & v_j \arrow[r, leftrightarrow, "\delta_j"] & v_j^{-1} \arrow[r, leftrightarrow, "(g^{-1})\rightarrow"] & v_r . \arrow[d, "(k)"] \\
 & & & \vdots
\end{tikzcd} \tag{3.4}$$

If v_r is minimal in $\boldsymbol{S}(c)$ and v_r does not support a unit circuit, then we repeat this process, performing the 3-gon move δ_r. If after this we obtain v_u minimal (not supporting a unit circuit), we perform δ_u. In this way, we are able to reverse v_u without first reversing v_i. Similarly, if (3.3) is a type LE(B) lateral escape, we have (3.4) above. If v_r is minimal, we perform δ_r. If we then obtain v_u minimal, we perform the 2-gon move γ_u and the 2-gon vertex v_i persists. The type LE(C) case is similar and is left as an exercise for the reader. In the type LE(D) case, it is clear that performing a 2-gon move for one of the $2k+1$ lateral escape 2-gon vertices reduces the number of such vertices by two, if $k>0$. Hence $2k$ of these moves removes all these 2-gons and also the turning vertex of the sub-graph. Finally, if

$$\begin{tikzcd}
 & v_u \arrow[dl, bend right] \arrow[dr, bend left, "(g)"] & \\
v_i \arrow[rr, Leftrightarrow, "(g)\rightarrow"] & & v_j
\end{tikzcd} \tag{3.5}$$

is a type LE(E) sub-graph and v_j is minimal in $\boldsymbol{S}(c)$, then δ_j removes both v_j and v_i from the shadow, $\boldsymbol{s}_u$, of v_u in $\boldsymbol{S}(c)$.

Using these observations, we are now in a position to prove the following key lemma:

Lemma 3.1. *Suppose that v_i is a 1-gon, weak 2-gon or a weak 3-gon vertex in $\boldsymbol{S}(c)$ with a lateral escape structure, $\boldsymbol{S}(v_i)$. Then $\boldsymbol{S}(v_i)$ defines a sequence of 1-gon, 2-gon and 3-gon moves, $\mathcal{T}(v_i)$, which remove v_i if v_i is a 2-gon vertex or reverse v_i if it is a 3-gon vertex. Moreover, the graph which results is the state graph, $\boldsymbol{S}(c')$, of a general position immersion, $c' : S^1 \to \Sigma$, in the homotopy class of c.*

Proof. Let V_i^0 denote the set of all vertices in $\boldsymbol{S}(v_i)$ which are minimal in $\boldsymbol{S}(c)$ and are not turning vertices in $\boldsymbol{S}(v_i)$. If $v_{k_m} \in V_i^0$, then we perform the elementary move associated with v_{k_m}. If v_{k_m} is a 1-gon or a 2-gon vertex, then we have clearly decreased the quantity $N(v_i)$ by one. If v_{k_m} is a 3-gon vertex, we claim that we decrease $N(v_i)$ by at least one. To see this, we note that if v_{k_m} is not the lateral escape vertex of a type LE(A), LE(B) or LE(C) sub-graph, then we simply remove $v_{k_m}^{-1}$ from the modified $\boldsymbol{S}(v_i)$ to obtain a lateral escape structure with one less vertex. If, on the other hand, v_{k_m} is such a lateral escape vertex, for instance we have

$$\begin{tikzcd} \vdots & \vdots \\ v_{k_{m-3}} \arrow[d] & v_{k_{m-1}} \arrow[d, "(g)"] \\ v_{k_{m-2}} & v_{k_m}, \arrow[l, leftrightarrow, "(h)\rightarrow"'] \end{tikzcd} \tag{3.6}$$

and δ_{k_m} acts to produce

$$\begin{tikzcd} \vdots & \vdots \\ v_{k_{m-2}} \arrow[d, "(g)"] & v_{k_{m-3}} \arrow[d, "(g^{-1}h)"] \\ v_{k_m}^{-1} \arrow[r, leftrightarrow, "(g^{-1})\rightarrow"] & v_{k_{m-1}} \arrow[d] \\ & \vdots \end{tikzcd} \tag{3.7}$$

then we remove the vertices $v_{k_m}^{-1}$ and $v_{k_{m-2}}$. The careful inductive definition of $\boldsymbol{S}(v_i)$ means that we retain the properties of a lateral escape structure throughout this process and $C(v_i)$ for the modified structure is two less than that for the original one. Similarly if v_{k_m} is the lateral escape vertex in a type LE(E) sub-graph. In the case that v_{k_m} is a lateral escape vertex in a type LE(D) sub-graph, the modified lateral escape structure also has complexity $C(v_i)$ two less than the original.

We repeat this process for each member of the set V_i^0, at the end of which we denote the resulting lateral escape structure by $\boldsymbol{S}^{(1)}(v_i)$. We then define the set V_i^1 in the analogous fashion to V_i^0 and continue in the same way until we have $C(v_i) = 0$. We finish by performing the elementary move for the vertex v_i and denote the overall sequence of moves by $\mathcal{T}(v_i)$. In the sequel we shall refer to $\mathcal{T}(v_i)$ as the sequence defined by *action on* $\boldsymbol{S}(v_i)$. $\square$

We note that at no stage in this procedure do we increase the number of double points of the immersed loop.

Suppose next that Γ is a spanning 2-gon for c. If Γ is $\pi_1(\Sigma)$-equivariant, then we may *cancel* it by equivariantly deforming one of its edges across the other. If Δ is not $\pi_1(\Sigma)$-equivariant, then we shall say that Δ is *cancellable* if we may equivariantly deform the lines in Λ until the 2-gon Δ is $\pi_1(\Sigma)$-equivariant without introducing any new 1-gons or 2-gons at any time. Once this is achieved we may cancel it as above. By [2], the cancellable 2-gons are precisely those which may be removed by cut-and-paste arguments. In particular, if $[c]$ is primitive in $\pi_1(\Sigma)$, then all 2-gons are cancellable. If $[c]$ is non-primitive, then we have an n-strand in Λ and $n-1$ non-cancellable 2-gons up to translation in $\pi_1(\Sigma)$. We may define cancellability for spanning 1-gons in a similar vein, noting, moreover, that all 1-gons are cancellable by [2]. When examining 3-gons, we have an analogous concept of *reversibility*. That is, we define a spanning 3-gon, Δ, for c to be *reversible* if we are able to equivariantly deform the members of Λ until Δ is $\pi_1(\Sigma)$-equivariant, after which we may reverse it in the standard fashion.

Lemma 3.2. *If a spanning* 1*-gon or* 2*-gon* Δ_j*, is cancellable, then we are able to transform* Δ_j *into an innermost* 2*-gon by a homotopy of* c*. Similarly, if* Γ_j *is a reversible spanning* 3*-gon, then we may transform it into an innermost* 3*-gon by a homotopy of* c*.*

Proof. We deal firstly with the case where Δ_j is a 2-gon, observing that if Δ_j is cancellable, then we may assume that it is $\pi_1(\Sigma)$-equivariant. We proceed with an explicit construction of a lateral escape structure for the vertex v_j associated to the disc Δ_j, which is contained in the shadow $\boldsymbol{s}_j$. We start by examining the set of sub-graphs of $\boldsymbol{S}(c)$ which are of one of the forms

$$\begin{tikzcd} & v_j \arrow[dl, bend right] \arrow[dr, bend left, "(g)"] & \\ v_{j_1} \arrow[rr, Leftrightarrow, "(g)\rightarrow"] & & v_{j_2}, \end{tikzcd} \tag{3.8}$$

$$\begin{array}{c} v_j \\ \Downarrow \\ v_{j_1} \end{array} \tag{3.9}$$

or

$$\begin{array}{c} v_j \\ \downarrow \\ v_{j_1}. \end{array} \tag{3.10}$$

The second of these, (3.9), refers to the special situation when v_{j_1} is a 1-gon and the third of these, (3.10), occurs precisely when we have neither (3.8) nor (3.9). It follows, in the case of (3.8), that v_{j_1} and v_{j_2} are 3-gon vertices and in the case of (3.10), the vertex v_{j_1} is either a 2-gon or a 1-gon vertex. In the former instance, we choose precisely one of the vertices v_{j_1} and v_{j_2} to be contained in a *level one sub-structure*, $\mathcal{S}^1(v_j)$, for v_j. Moreover, we make this choice, where necessary, subject to one important constraint, which we describe next. In particular, if in addition to a sub-graph (3.8), we have a sub-graph

$$\begin{tikzcd} & v_j \arrow[dl, bend right] \arrow[dr, bend left, "(h)"] & \\ v_{j_3} \arrow[rr, Leftrightarrow, "(h)\rightarrow"] & & v_{j_4}, \end{tikzcd} \tag{3.11}$$

where v_{j_3} is neither v_{j_1} nor v_{j_2} and suppose, moreover, that

$$\begin{array}{c} v_{j_1} \\ \downarrow^{(h)} \\ v_{j_4} \end{array} \tag{3.12}$$

and

$$\begin{array}{c} v_{j_3} \\ \downarrow^{(g)} \\ v_{j_2}. \end{array} \tag{3.13}$$

Then if $\mathcal{S}^1(v_j)$ contains the sub-graph

$$\begin{array}{c} v_j \\ \downarrow \\ v_{j_1}, \end{array} \tag{3.14}$$

then it must also contain the sub-graph

$$\begin{array}{c} v_j \\ \downarrow \\ v_{j_4}. \end{array} \tag{3.15}$$

In the event that we have the sub-graph (3.9) where v_{j_1} is either a 1-gon and v_{j_2} is a 2-gon, we define $\mathcal{S}^1(v_j)$ to contain the vertex v_{j_1} and the vertical edge connecting it to v_j. Finally, if we have the third situation, (3.10), then we define $\mathcal{S}^1(v_j)$ to contain this sub-graph.

We note that according to this construction, no two vertices in $\mathcal{S}^1(v_j)$ are horizontally related in $\boldsymbol{S}(c)$ and hence this sub-graph is compatible in the sense of Definition 2.12.

Suppose next that we have constructed a sub-structure for each level less than or equal to some $k \in \mathbb{N}$ for the vertex v_j. Then we define a level $k+1$ *sub-structure*, $\mathcal{S}^{k+1}(v_j)$, for v_j by repeating the above process for each vertex which is minimal in $\mathcal{S}^k(v_j)$. In this more general situation, we must examine sub-graphs of the form

$$\begin{tikzcd} & v_s \arrow[dl, bend right] \arrow[dr, bend left, "(g)"] & \\ v_{s_1} \arrow[rr, Leftrightarrow, "(g)\to"] & & v_{s_2}, \end{tikzcd} \tag{3.16}$$

or

$$\begin{array}{c} v_s \\ \downarrow \\ v_{s_1}, \end{array} \tag{3.17}$$

where v_s may be a 1-gon or a 3-gon vertex. If v_s is a 1-gon, then either v_{s_1} is a 1-gon vertex and v_{s_2} is a 2-gon vertex in (3.16) or v_{s_1} is a 1-gon vertex in (3.17). We note that the latter describes the situation in which a representative 1-gon disc for v_s interior includes a representative 1-gon disc for v_{s_1}. Given a sub-graph of the form (3.16), we define $\mathcal{S}^{k+1}(v_j)$ to contain 1-gon vertex v_{j_1}. Similarly, given the sub-graph (3.17), we define $\mathcal{S}^{k+1}(v_j)$ to contain it. If v_s is a 3-gon vertex, then v_{s_1} may be any of a 1-gon, 2-gon or a 3-gon vertex and we have a sub-graph of the form (3.17) in $\mathcal{S}^{k+1}(v_j)$.

It it is easy to check that this process terminates at some finite level, n, with a lateral escape structure, $\mathcal{S}^n(v_j)$, for the vertex v_j which lies in the shadow $\boldsymbol{s}_j$. $\square$

In the second key lemma, we show that all cancellable 2-gons and 1-gons (the latter are invariably cancellable) are weak according to our definition, and so can be removed using an appropriate sequence of the elementary homotopies. To this end, suppose that $c_1 : S^1 \to \Sigma$ is a general position immersion which is homotopic to c and that c_1 has the minimal intersection property. We then have the following fairly basic combinatorial measure of distance between the curves c and c_1, given by

$$d(c, c_1) = (D, E)(c, c_1), \tag{3.18}$$

where D measures the number of cancellable 2-gon vertices in $\boldsymbol{S}(c)$ and E measures the number of (cancellable) 1-gon vertices in $\boldsymbol{S}(c)$. We note that the condition $d(c,c_1)=(0,0)$ is insufficient to ensure that c is ambient isotopic to c_1, since we may still have very different configurations of spanning 3-gons for the loops c and c_1. We therefore refine this distance measure later on.

Lemma 3.3. *Suppose that c and c_1 are immersed loops as above for which $d(c,c_1)\neq(0,0)$. Then given a 1-gon or cancellable 2-gon vertex, v_i, in $\boldsymbol{S}(c)$, there exists a lateral escape structure for v_i, action on which produces an immersion, $c':S^1\to\Sigma$, having $d(c',c_1)=(0,0)$.*

Proof. Suppose firstly that $\Lambda^{(k)}$ is a finite subset of lines in Λ. We shall define a *configuration disc*, $D(\Lambda^{(k)})$, for $\Lambda^{(k)}$ to be an embedded 2-disc in $\widetilde{\Sigma}$ which has the following properties.

1. If l and m are distinct lines in $\Lambda^{(k)}$ which do not share an infinite cyclic stabilizer sub-group in $\pi_1(\Sigma)$, then the points of $l\cap m$ are all contained in $\operatorname{Int} D(\Lambda^{(k)})$.
2. If two distinct intersecting lines, l and m, in $\Lambda^{(k)}$ share an infinite cyclic sub-group, $\langle h\rangle$, in $\pi_1(\Sigma)$, then $\operatorname{Int} D(\Lambda^{(k)})$ contains precisely one representative of each of the $\langle h\rangle$-orbits in $l\cap m$.
3. Each line λ in $\Lambda^{(k)}$ meets $D(\Lambda^{(k)})$ in a single line segment.
4. The circle $\partial D(\Lambda^{(k)})$ is transverse to the members of $\Lambda^{(k)}$.

Given a configuration disc, $D(\Lambda^{(k)})$, for $\Lambda^{(k)}$, we define the associated *configuration circle*, $C(\Lambda^{(k)})$, to be the boundary, $\partial D(\Lambda^{(k)})$, of the configuration disc. Moreover, we define a *configuration sequence*, $S(\Lambda^{(k)})$, for $\Lambda^{(k)}$ to be a sequence whose members are the points of the set

$$C\bigl(\Lambda^{(k)}\bigr)\cap\Bigl(\bigcup\Lambda^{(k)}\Bigr), \tag{3.19}$$

in a clockwise ordering on the configuration circle, $C(\Lambda^{(k)})$. We say that a point, p, in a configuration sequence, $S(\Lambda^{(k)})$, is *derived from* the line $\lambda\in\Lambda^{(k)}$ if it lies in the set $C(\Lambda^{(k)})\cap\lambda$. Clearly, we have two derived points in $S(\Lambda^{(k)})$ for each line $\lambda\in\Lambda^{(k)}$. In order to distinguish these two points, we shall say that a point, p, in $C(\Lambda^{(k)})$ is *positive* if in moving clockwise in $S(\Lambda^{(k)})$ through x we pass from the positive to the negative side of the line λ. Otherwise, we shall say that p is *negative*. We see immediately that the two points of $\lambda\cap C(\Lambda^{(k)})$ form a pair, one member of which is positive and the other of which is negative. We may then describe $S(\Lambda^{(k)})$ unambiguously by a sequence of lines in $\Lambda^{(k)}$ in which each member of $\Lambda^{(k)}$ appears twice, once marked positive and once marked negative, and the ordering of the underlying points in $C(\Lambda^{(k)})$ is that of a configuration sequence for $\Lambda^{(k)}$. We shall call this sequence a *configuration pattern* for $\Lambda^{(k)}$. The key feature to observe here is that if no two lines in $\Lambda^{(k)}$ share an infinite cyclic stabilizer, then a configuration pattern for $\Lambda^{(k)}$ corresponds under the bijection $\Phi:\Lambda\to\Lambda'$ up to cyclic permutation to a configuration pattern for the set, $\Lambda^{(k)\prime}$, of Φ-corresponding lines in Λ'. On the other hand, if two distinct lines, l and m, in $\Lambda^{(k)}$ share an infinite cyclic

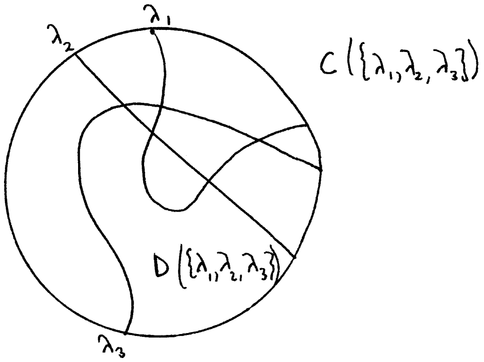

Fig. 11. A configuration disc in $\widetilde{\Sigma}$.

stabilizer sub-group in $\pi_1(\Sigma)$, then $m = hl$ for some $h \in \pi_1(\Sigma)$, since these lines are lifts of the same immersed loop, c, in Σ. It follows that l and m lie in an n-strand for some $n \in \mathbb{N}$, allowing us a choice of $n-1$ configuration discs with distinct configuration patterns up to cyclic permutation. It is nonetheless possible in this situation to choose a pair of configuration discs, one for $\Lambda^{(k)}$ and one for $\Lambda'^{(k)}$, for which the configuration patterns correspond under Φ. In the sequel, we assume that such a choice has been made. We note moreover, that this form of rigidity, albeit rather weak, allows us to list the possible 1-complexes, $\Lambda'^{(k)} \cap D(\Lambda'^{(k)})$, given a configuration pattern for $\Lambda^{(k)}$. We start by constructing a lateral escape structure for a single cancellable 2-gon vertex, v_i, in $\boldsymbol{S}(c)$. We note, moreover, that since v_i is cancellable, then by Lemma 3.2, we may homotop c to an immersed, general position loop, $d : S^1 \to \Sigma$, whilst converting v_i to a minimal 2-gon vertex in $\boldsymbol{S}(c)$. We go on to show that the situation is similar if v_i is a 1-gon or a reversible 3-gon. Prior to making our construction, we introduce one more concept. In particular, let us denote by H the homotopy which converts c to the map d above, where $H(c,0) = c$ and $H(c,1) = d$. Suppose, moreover, that $H(c,t)$ is a general position immersion for all but finitely many values of t in $(0,1)$ and H introduces no 1-gons or 2-gons. Suppose, in addition, that none of the maps, $H(c,t),\ t \in (0,1)$, has a non-transverse triple point, that is, a triple point where two of the lifts are tangential. Then we shall say that H is a *simple homotopy*. We note that the last condition is equivalent to the requirement that a 2-gon may be cancelled in the course of the homotopy only if it is first made innermost. We claim that given an arbitrary homotopy, H, carrying the map c to the map d, there is always a simple homotopy which performs the same function. To see this, we note that we may achieve the first condition by perturbation. The second condition follows from the definition of cancellability. In particular, any new 2-gon or 1-gon would get removed later on in the homotopy. Hence we may continuously deform the homotopy so as to prevent its introduction. It remains only to examine the third condition. In particular, we are concerned with the ways in which non-transverse triple points might arise. We note indeed that the case of three tangential curve segments may be eliminated by perturbation, leaving us with the case where two of the curve segments are tangential and the third is transverse to both of these at a point of intersection. We illustrate this in Fig. 12. The only instance where

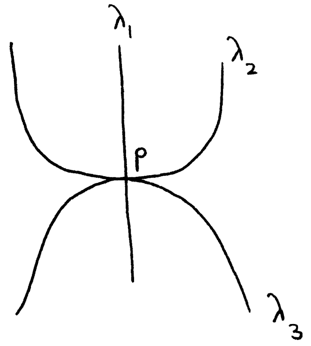

Fig. 12.

this is not easily removable by perturbation is when we are cancelling a 2-gon which is non-innermost. Then a traversing arc will give this sort of intersection. We have, however, that a cancellable 2-gon may be homotoped to be innermost. Hence we need only compose the two homotopies. Repeating this as many times as is necessary gives us the desired conclusion. Hence we work from now on with simple homotopies.

We work using a series of local observations, starting by examining the sub-graphs of $\boldsymbol{S}(c)$ whose forms are amongst the following:

$$\begin{tikzcd} & v_i \arrow[dl, bend right] \arrow[dr, "(g)"] & \\ v_{i_1} \arrow[rr, leftrightarrow, "(g)\rightarrow"] & & v_{i_2}, \end{tikzcd} \tag{3.20}$$

$$\begin{array}{c} v_i \\ \Downarrow \\ v_{i_1}, \end{array} \tag{3.21}$$

and, in the absence of the above, given vertices v_i and v_{i_1},

$$\begin{array}{c} v_i \\ \downarrow \\ v_{i_1}. \end{array} \tag{3.22}$$

In the event of (3.20), v_{i_1} and v_{i_2} are necessarily both 3-gons. In the case of (3.21), v_{i_1} is a 1-gon and v_{i_2} is a 2-gon and in (3.22), v_{i_1} may be either a 1-gon or a 2-gon. Suppose then that we have the graph (3.20) where v_{i_1} and v_{i_2} are both 3-gons. We consider representative discs, Γ_i, Δ_{i_1} and Δ_{i_2} for these vertices, where $\Delta_{i_1} \cup \Delta_{i_2} = \Gamma_i$. These discs have edges lying in a total of three distinct lines, λ_1, λ_2 and λ_3 covering c. We write $\Lambda^{(3)}$ for the set $\{\lambda_1, \lambda_2, \lambda_3\}$ and focus on the configuration disc $D(\Lambda^{(3)})$. Now homotoping the map c to the map d by a simple homotopy, H, converts the 2-gon Γ_i to an innermost 2-gon and leaves the configuration pattern for the collection $\Lambda^{(3)}$ fixed. Using this information, we can draw some conclusions about the effect of H on Δ_{i_1} and Δ_{i_2}. In particular, we must have one of the following effects under the homotopy H.

(1) H reverses Δ_{i_1},
(2) H reverses Δ_{i_2},

(3) H reverses neither Δ_{i_1} nor Δ_{i_2}, but Δ_{i_2} is the partner of some Δ_{j_1} in a 2-gon Γ_j, with edge segments in the lines λ_2 and λ_3. Moreover, H reverses Δ_{j_1} and then later cancels Γ_j. We note that since H is simple, there may be a number of moves occurring in between the reversal of Δ_{j_1} an the cancellation of Γ_j.

In the event of situation 1, we define the *level one sub-structure*, $\boldsymbol{S}^1(v_i)$, to contain the subgraph (3.22). Symmetrically, if we have the situation 2, then we define $\boldsymbol{S}^1(v_i)$ to contain the sub-graph

$$\begin{tikzcd} v_i \arrow[d, "(g)"] \\ v_{i_2}. \end{tikzcd} \tag{3.23}$$

Finally, in the event of situation 3 we define $\boldsymbol{S}^1(v_i)$ to contain the type LE(C) lateral escape sub-graph,

$$\begin{tikzcd} v_i \arrow[d] & v_j \arrow[d, "(h)"] \\ v_{i_1} \arrow[r, leftrightarrow, "(h)\rightarrow"] & v_{j_1}. \end{tikzcd} \tag{3.24}$$

Suppose next that we have the sub-graph (3.22) and that v_{i_1} is also a 2-gon vertex. The simple homotopy H must have one of the following effects.

(1) H cancels the 2-gon Γ_{i_1},

(2) H does not cancel Γ_{i_1} and we have a type LE(B) lateral escape sub-graph,

$$\begin{tikzcd} v_i \arrow[d] & v_{i_s} \arrow[d, "(h)"] \\ v_{i_1} \arrow[r, leftrightarrow, "(h)\rightarrow"] & v_{i_4}, \end{tikzcd} \tag{3.25}$$

where v_{i_3} and v_{i_4} are 3-gon vertices. Moreover, H acts to reverse the 3-gon Δ_{i_3} and then later reverses either the 3-gon Δ_{i_4}, or its partner 3-gon in Γ_i.

In the event of 1, we define the level one sub-structure, $\boldsymbol{S}^1(v_i)$, to contain the sub-graph (3.22). If we have the situation 2, we define $\boldsymbol{S}^1(v_i)$ to contain the type LE(B) sub-graph (3.25).

Suppose next that we have the situation (3.22) where v_{i_1} is a 1-gon vertex. We then have two possibilities for the homotopy H.

(1) H removes the 1-gon Z_{i_1},

(2) H reverses the 3-gon, Δ_j, which is vertex adjacent to Z_{i_1} and edge included by Γ_i, thereby converting Z_{i_1} into a 2-gon. The homotopy H then cancels this 2-gon.

If we have the situation 1, then we define $\boldsymbol{S}^1(v_i)$ to contain the sub-graph (3.22). If, on the other hand, we have the situation 2, then we define $\boldsymbol{S}^1(v_i)$ to contain the sub-graph

$$\begin{tikzcd} v_i \arrow[d] & \\ v_{i_1} \arrow[r, leftrightarrow, "(g)\rightarrow"] & v_j. \end{tikzcd} \tag{3.26}$$

Finally, we may have the situation (3.21), where v_{i_1} is a 1-gon. In this very special case, the homotopy H must act to eliminate the 1-gon v_{i_1}, and we define $\boldsymbol{S}^1(v_i)$ to contain the sub-graph (3.21).

We next turn to the case where the vertex v_i is a 1-gon. If v_i is not minimal in $\boldsymbol{S}(c)$, then it must split, either according to (3.20) into a 2-gon and a 3-gon, or it must lie in a sub-graph of the form (3.22), where v_{i_1} is a 1-gon vertex. In the former case, we have two possibilities for H.

(1) H cancels the sub-2-gon, Γ_{i_1},

(2) H reverses the sub-3-gon, Δ_{i_2}.

If we have the situation of 1, then we define $\boldsymbol{S}^1(v_i)$ to contain the sub-graph (3.22). If we have that of 2, then we define $\boldsymbol{S}^1(v_i)$ to contain the sub-graph (3.23).

In the latter case, where we have the sub-graph (3.22), the homotopy H must act to remove the 1-gon Z_{i_1}, and we define $\boldsymbol{S}^1(v_i)$ to contain the sub-graph (3.22).

Hence, given a 1-gon or a cancellable 2-gon vertex, v_i, in $\boldsymbol{S}(c)$, we build up a level one sub-structure, $\boldsymbol{S}^1(v_i)$. We note, moreover, that this sub-graph will satisfy the compatibility criteria for a lateral escape structure. We extend this sub-graph inductively to obtain the sought-after lateral escape structure. In particular, suppose henceforth that we have defined a level k sub-structure, $\boldsymbol{S}^k(v_i)$, for v_i for some $k \in \mathbb{N}$. We construct a level $k+1$ sub-structure, $\boldsymbol{S}^{k+1}(v_i)$, from $\boldsymbol{S}^k(v_i)$ by examining the set of vertices which are minimal in $\boldsymbol{S}^k(v_i)$ and are not turning vertices of any lateral escape sub-graphs in $\boldsymbol{S}^k(v_i)$. Indeed, if v_{i_k} is some such vertex which is either a 1-gon or a 2-gon, then we extend $\boldsymbol{S}^k(v_i)$ by adding a level one sub-structure for v_{i_k} in the same manner as before. It remains then to consider the case where v_{i_k} is a 3-gon vertex, where we note that the simple homotopy H acts, by definition, to reverse the underlying 3-gon disc, Δ_{i_k}. We proceed by examining sub-graphs of the form

$$\begin{array}{c} v_{i_k} \\ \downarrow \\ v_{i_{k+1}}, \end{array} \tag{3.27}$$

noting that here we have no analogue of the sub-graph (3.20). The vertex $v_{i_{k+1}}$ may be any of a 1-gon, 2-gon or a 3-gon vertex. In the case that $v_{i_{k+1}}$ is a 1-gon, the homotopy H must eliminate the underlying 1-gon disc, $Z_{i_{k+1}}$, and we define $\boldsymbol{S}^{k+1}(v_i)$ to contain the sub-graph (3.27). If $v_{i_{k+1}}$ is a 2-gon vertex, then we have the following possibilities.

(1) H cancels the 2-gon disc, $\Delta_{i_{k+1}}$,

(2) v_{i_k} and $v_{i_{k+1}}$ lie in a type LE(B) lateral escape sub-graph,

$$\begin{tikzcd} v_{i_k} \arrow[d] & \\ v_{i_{k+1}} & v_{i_{k+2}}, \arrow[l, "(g)\rightarrow"'] \end{tikzcd} \tag{3.28}$$

in $\boldsymbol{S}(c)$. Moreover, H reverses the lateral escape 3-gon $\Delta_{i_{k+2}}$ and subsequently reverses the resulting sub-3-gon, $\Delta_{i_{k+1}}$, which replaces the 2-gon $\Gamma_{i_{k+1}}$.

In the first instance, we define $\boldsymbol{S}^{k+1}(v_i)$ to contain the sub-graph (3.27). If we have the situation of 2 on the other hand, then we define $\boldsymbol{S}^{k+1}(v_i)$ to contain the sub-graph (3.28).

Suppose next that $v_{i_{k+1}}$ is a 3-gon. The homotopy H must act in one of the following ways.

(1) H reverses the 3-gon $\Delta_{i_{k+1}}$,

(2) v_{i_k} and $v_{i_{k+1}}$ lie in a type LE(A) lateral escape sub-graph,

$$\begin{tikzcd} v_{i_k} \arrow[d] & & v_{i_{k+3}} \arrow[d, "(g)"] \\ v_{i_{k+2}} \arrow[rr, leftrightarrow, "(g)\rightarrow"] & & v_{i_{k+4}}, \end{tikzcd} \tag{3.29}$$

where H reverses the 3-gon $\Delta_{i_{k+4}}$ and then subsequently reverses the 3-gon $\Delta_{i_{k+3}}$, followed by Δ_{i_k}.

In the first case, we define $\boldsymbol{S}^{k+1}(v_i)$ to contain the sub-graph (3.27) and in the second, we define $\boldsymbol{S}^{k+1}(v_i)$ to contain the lateral escape sub-graph (3.29).

We note that the level $k+1$ sub-structure, $\boldsymbol{S}^{k+1}(v_i)$, defined in this way satisfies the compatibility criteria for a lateral escape structure. Moreover, there must exist some $n \in \mathbb{N}$ for which $\boldsymbol{S}^n(v_i)$ is a lateral escape structure, since there exists a simple homotopy H. □

We let $\mathcal{T}(c)$ denote the sequence of 1-gon, 2-gon and 3-gon moves derived from action on this lateral escape structure. Applying this sequence of moves to the curve c removes the vertex v_i, giving us a new immersed loop, c'. We continue by locating a 1-gon or cancellable 2-gon for c' and applying the same procedure. Eventually we are left with a loop which has no 1-gons or cancellable 2-gons and we have proved Theorem 1.1.

Suppose, in the sequel, that c and c' are homotopic, general position immersions with the minimal intersection property. As mentioned earlier, these need not be ambient isotopic. We therefore develop our existing techniques further by designing a combinatorial algorithm for homotoping c until it is ambient isotopic with c'. We do this starting with the "easy" case where c (and therefore also c') carries a primitive element of $\pi_1(\Sigma)$. In this circumstance we have no 1-gons or 2-gons. In particular, we may bijectively pair a 3-gon vertex in $\boldsymbol{S}(c)$ with a 3-gon vertex in $\boldsymbol{S}(c')$, where they both have corresponding underlying spanning 3-gon discs. We then focus on those vertices which reverse correspond to their counterparts under this pairing. Using this, we define a new notion of *distance*, $\Delta(c,c')$, between c and c', which is the number of pairs of reverse corresponding 3-gon vertices for $\boldsymbol{S}(c)$ and $\boldsymbol{S}(c')$. Clearly,

$$\Delta\big(c,c'\big) = 0 \tag{3.30}$$

if and only if c is ambient isotopic to c'. Suppose then that $\Delta(c,c') > 0$. It follows, since c has no spanning 1-gons or 2-gons that there must exist some minimal 3-gon vertex, v_i, in $\boldsymbol{S}(c)$ which reverse corresponds to its counterpart, v_i', in $\boldsymbol{S}(c')$. To see this, we recall the argument of Lemma 3.3 above. In particular, since $\Delta(c,c')$ is non-zero, there must exist some 3-gon vertex, v_j, in $\boldsymbol{S}(c)$ which reverse corresponds to some 3-gon vertex, v_j', in $\boldsymbol{S}(c')$. Hence there must exist a simple homotopy, H, which makes v_j minimal in $\boldsymbol{S}(c)$ and this homotopy must reverse some innermost 3-gon for c, as required. We are therefore able to reduce $\Delta(c,c')$ by one through performing this 3-gon move. We repeat this process until

$\Delta(c, c')$ is zero. We note that since v_j is reversible, it may be made minimal by action on a lateral escape structure, $\boldsymbol{S}(v_j)$, determined as in Lemma 3.3 above. This action reduces $\Delta(c, c')$ by at least one, since we define our lateral escape structure in terms of the simple homotopy H.

The problem becomes more intricate if we allow c to carry a non-primitive element of $\pi_1(\Sigma)$. Indeed, we must take into account non-cancellable 2-gons in n-strands. In particular, suppose that c lifts to an n-strand consisting of lines, λ, $g\lambda, \ldots, g^{n-1}\lambda$, where $n \in \mathbb{N}$, $g^n\lambda = \lambda$, and that $h\lambda$ is some other lift of c which crosses one of the 2-gons for this n-strand and is not a member of the n-strand itself. We denote the collection of lines, $\{\lambda, g\lambda, h\lambda\}$, by $\Lambda^{(g,h)}$ and work in a configuration disc, $D(\Lambda^{(g,h)})$, for $\Lambda^{(g,h)}$. As mentioned earlier, we select a configuration disc, $D(\Lambda'^{(g,h)})$, for the set of Φ-corresponding lines, $\Lambda'^{(g,h)} = \{\lambda', g\lambda', h\lambda'\}$, with a corresponding configuration pattern. That is, the clock-wise sequence of lines and associated signs encountered on the boundary of $D(\Lambda'^{(g,h)})$ corresponds to a cyclic permutation of that for $D(\Lambda^{(g,h)})$. We note that the line $h\lambda$ may only meet a single 2-gon with boundary segments in λ and $g\lambda$ and that it may only split this 2-gon into a pair of sub-3-gons. Otherwise it is easy to see that it would have to belong to the n-strand itself, contrary to our assumption. As in the case of the primitive loop, we can establish a bijection between the 3-gon vertices of $\boldsymbol{S}(c)$ and those of $\boldsymbol{S}(c')$ based on the relation of correspondence. It may be the case, however, that we need to perform a succession of 3-gon moves, deforming $h\lambda$ over more than one double point for the same pair of lines to make the picture in $D(\Lambda^{(g,h)})$ look like that in $D(\Lambda'^{(g,h)})$. We therefore need to refine our measure of distance. In particular, we define $\Delta^*(c, c')$, to be given by the lexicographically ordered pair,

$$(D, \Delta)\bigl(c, c'\bigr), \tag{3.31}$$

where Δ is as above. In order to define the quantity S, we return to examine the discs $D(\Lambda^{(g,h)})$ and $D(\Lambda'^{(g,h)})$ and suppose that the 1-complexes for $\Lambda^{(g,h)}$ and $\Lambda'^{(g,h)}$ in their respective configuration discs are non-isomorphic. We then define $D_{g,h}$ to be the maximum number of whole 2-gons with boundary segments in λ and $g\lambda$ across which $h\lambda$ must be deformed in $D(\Lambda^{(g,h)})$, for the 1-complex associated with $\Lambda^{(g,h)}$ in $D(\Lambda^{(g,h)})$ to be isomorphic with that for $\Lambda'^{(g,h)}$ in $D(\Lambda'^{(g,h)})$. This maximum will be formed over at most two numbers. To see this, we refer the reader to Fig. 13, in which we illustrate a situation where this quantity is two, whilst a shorter "zero-length" path still exists. We define these quantities for each set of the form $\Lambda^{(g,h)}$, where the line λ is fixed and we examine precisely one representative per $\pi_1(\Sigma)$-orbit of ordered pair (g, h). We then define $D(c, c')$ by summing the terms $D_{g,h}$.

Lemma 3.4. *Suppose that c and c' are homotopic immersions with the minimal intersection property and that $\Delta^*(c, c')$ is non-zero. Then there exists a reversible 3-gon, action on which decreases $\Delta^*(c, c')$ by at least one.*

Proof. If $D(c, c') = 0$, then we proceed precisely as in the case of a primitive loop described above. If, on the other hand, $D(c, c')$ is non-zero, then some term, $D_{g,h}$, is

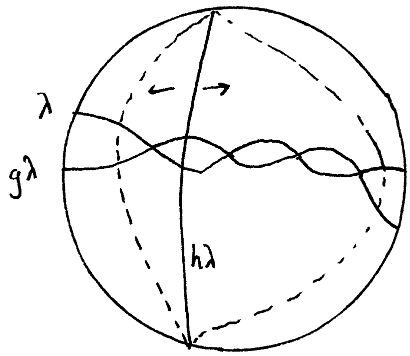

Fig. 13.

non-zero and from examining the configuration disc for these three lines, it is easy to see that there must exist a reversible 3-gon, v_j. Moreover, the simple homotopy H may be described in the vicinity of v_j by action on some lateral escape structure, $\boldsymbol{S}(v_j)$, determined as in Lemma 3.3 above. Action on this lateral escape structure must, by its construction, reduce either $D(c, c')$ or $\Delta(c, c')$ by at least one. $\square$

We continue this process until $\Delta^*(c, c') = (0, 0)$ and the modified loop c is ambient isotopic to c'. Writing $\mathcal{T}(c, c')$ to denote a sequence of 3-gon moves defined as above, and combining this with Theorem 1.1, we have proved Theorem 1.2.

We may summarize the achievements of this paper as follows. In particular, given an immersed loop, c, in a closed, orientable surface, Σ, we can construct the state graph, $\boldsymbol{S}(c)$, of c, from examining the arrangement of its spanning discs in the universal covering plane $\widetilde{\Sigma}$. We then examine the 1-gon and 2-gon vertices of $\boldsymbol{S}(c)$ and construct a sequence of lateral escape structures, successive action on whose terms eliminates all the weak 2-gons and 1-gons in $\boldsymbol{S}(c)$. Moreover, the existence of this sequence is guaranteed and we use it to replace c by an immersion, c', which is homotopic to c and has the minimal number of self-intersections of any general position loop in its homotopy class. We extend this technology to produce an algorithm which converts a minimally self-intersecting general position immersed loop, c, to one which is ambient isotopic to any other specified homotopic general position immersed loop, d. This amounts to locating a sequence of lateral escape structures, successive action on which reverses the “right” 3-gons and so moves the immersion closer to d. We note, moreover, that the construction of the state graph and the search for lateral escape structures should be achievable by hand in many cases and relatively straightforward to implement by computer.

Finally, we note that this technology should transfer to the situation where we have two immersed curves, c and d, in Σ in general position, where we wish to minimize the number of points in $c \cap d$. To do this, we would need to set up a state graph, $\boldsymbol{S}(c, d)$, the vertices of which are derived from spanning 2-gons and 3-gons with edges in both lifts of c and lifts of d. By making the analogous constructions, we should then have an algorithm which minimizes $|c \cap d|$ by a sequence of 2-gon and 3-gon moves.

Acknowledgements

I would like to thank Joel Hass and Peter Scott for helpful conversations.

References

[1] M. Cohen, M. Lustig, Paths of geodesics and geometric intersection numbers: 1, in: S.M. Gersten, J.R. Stallings (Eds.), Combinatorial Group Theory and Topology, Ann. of Math. Stud., Vol. 111, Princeton Univ. Press, Princeton, NJ, 1987, pp. 479–500.
[2] M. Dehn, Papers on Group Theory and Topology, Springer, Berlin, 1985. Transl. J. Stillwell.
[3] J. Hass, P. Scott, Intersection of curves on surfaces, Israel J. Math. 51 (1–2) (1985) 90–120.
[4] J. Hass, P. Scott, Shortening curves on surfaces, Topology 33 (1) (1994) 25–43.
[5] J. Paterson, Football regions for π_{i_1}-injective immersed surfaces in 3-manifolds, Preprint.